

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 02-136900

(43)Date of publication of application : 25.05.1990

(51)Int.Cl.

G10L 9/20
G06F 15/40
G09F 25/00
H01L 27/148

(21)Application number : 63-291586

(71)Applicant : NEC CORP

(22)Date of filing : 18.11.1988

(72)Inventor : TAKAHASHI TSUNESUKE

(54) VOICE TEXT RETRIEVING DEVICE

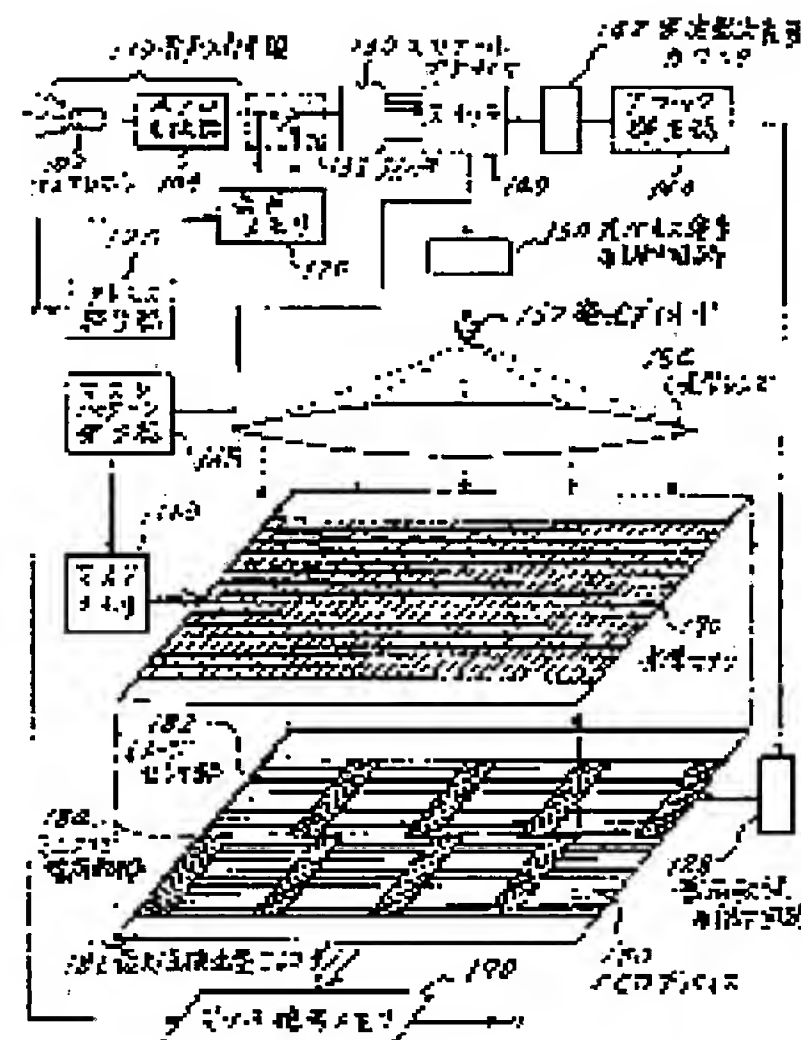
(57)Abstract:

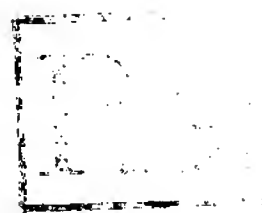
PURPOSE: To prevent a decline in voice pattern recognition rate resulting from the up-down and left-right variation and concentration variation of voice spectrogram at every hour by providing a spectrum analyzing means, optical pulse generating means, optical mask means, and matching CCD device.

CONSTITUTION: Analog signal waveforms of each frequency of voices obtained by means of a spectrum analyzer 130 are scanned in the order of the frequency at every hour and converted into optical pulse strings and the whole surface of an optical mask 170 is irradiated with the optical pulses so that the analog signal waveforms of all frequencies of registered voices can be compared with the components at every hour. The sum of the optical pulses passed through the optical mask 170 are taken on a CCD device 180 at every hour and similarity between inputted voices and each registered voice is found by processing the results by a dynamic programming method. Then class

discrimination is performed by means of a maximum value detection type encoder 186.

Therefore, a decline in voice recognition rate resulting from the hourly change and variation in pitch and loudness of inputted voices can be prevented.





2

⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A)

平2-136900

⑬ Int. Cl.⁵

識別記号

庁内整理番号

⑭ 公開 平成2年(1990)5月25日

G 10 L 9/20
G 06 F 15/40
G 09 F 25/00
H 01 L 27/148

3 0 1 B
5 3 0 H

8842-5D
7313-5B
6422-5C

7377-5F H 01 L 27/14

B

審査請求 未請求 請求項の数 3 (全10頁)

⑮ 発明の名称 音声テキスト検索装置

⑯ 特 願 昭63-291586

⑰ 出 願 昭63(1988)11月18日

⑱ 発 明 者 高 橋 恒 介 東京都港区芝5丁目33番1号 日本電気株式会社内

⑲ 出 願 人 日本電気株式会社 東京都港区芝5丁目33番1号

⑳ 代 理 人 弁理士 内 原 晋

明 細 書

発 明 の 名 称

音声テキスト検索装置

特 許 請 求 の 範 囲

(1) 音声テキスト及び検索用単語音声周波数別のアナログ信号波形に変換するスペクトル分析手段と、変換された前記アナログ信号波形の各時刻でのアナログ信号振幅を周波数順に光パルス信号に変換する光パルス発生手段と、前記検索用単語音声を重み係数の光マスクパターンに変換して保持すると共に前記光パルス信号を受ける光学マスク手段と、前記光学マスク手段を通過した前記光パルス信号を検出し、前記アナログ信号波形のマッチング処理をし、マッチした単語音声のクラス判別を行うマッチングCCDデバイスとを備えたことを特徴とする音声テキスト検索装置。

(2) 特許請求の範囲第(1)項に記載の音声テキスト検索装置において、前記マッチングCCD

Dデバイスが前記光学マスク手段を通過した前記光パルス信号を検出し、周波数毎の検出信号を積算して転送するイメージセンサ手段と、前記イメージセンサ手段に結合して各時刻における前記イメージセンサ手段の各段の出力とそれ以前の出力との中から最大値を選び、前記最大値と前段からの転送信号とを比較して最小振幅の方を選択し、次段への転送信号とするミニマクス電荷転送手段と、複数の前記ミニマクス電荷転送手段の出力の中の最大値を検出し、その位置をエンコードする最大検出エンコード手段とを備えたことを特徴とする音声テキスト検索装置。

(3) 特許請求の範囲第(1)項に記載の音声テキスト検索装置において、前記音声テキストをデジタル信号波形に変換して記憶する音声テキスト記憶手段と、前記検索用単語音声と、前記音声テキスト記憶手段のレコードアドレスと前記マッチングCCDデバイスの出力するクラスコードとでアクセス番地の決まるマッチビット記憶手段とを備えたことを特徴とする音声テキスト検索

装置。

発明の詳細な説明

〔産業上の利用分野〕

本発明は音声テキスト検索装置に関し、特に音声テキストや録音テキストの内容検索に関する。

〔従来の技術〕

従来、ラジオやテレビ等でアナウンスされる音声テキストはホットな知識情報を多く含んでいるが、それを聞いて必要な部分を録音するとか、メモを取るとかは人間にとって簡単にできることではあるが楽な仕事ではない。それだけに機械にやらせればよいと考えられてきた。特に外国語の音声テキストについてはキーポイントとなる単語音声の有り無しを高速に判断する必要がある、高速の音声テキスト検索装置の実現への期待が大きい。しかしながら、従来は音声認識機械によって音声テキストの全てを認識し、且つコード化し、その後で検索を行えるようにしようしていた。

〔発明が解決しようとする課題〕

れるようになったが、声楽曲のように揺れる音声の音の高さや揺れに対しては余り許容度がなかった。

従来の柔軟性の少ない音声認識機械では、音声テキストの各部の音声の高さや強さが録音時の単語音声の高さや強さに一致していないとマッチングが悪くなった。また、音声を高さや強さを変えて別々に登録しておかないと分類が正しく行われないような認識装置では実用度が低くなる。このようなこともあって、ラジオやテレビからの生放送の音声テキストの検索が実現されなかった。

ここで音声テキストの高さや強さの揺れについて以下に詳細に説明する。まず、単語音声を認識するためには、音声信号が電気信号に変換され、高速フーリエ変換器などのスペクトラムアナライザによって周波数毎のアナログ電気信号波形に分解され、音声パターンになる。第2図は音声パターンの説明図である。横線が時間軸201、縦線が周波数軸202である。観測される音声スペクトログラム203は各時刻での各周波数における電気

上述した従来の音声テキスト検索装置は、音声認識機械によって検索を行えるようにしようとしたが、このような考え方では音声認識技術が十分に高くなしないと、音声テキストの検索は有り得ないことであった。また、一般に高性能な音声認識機械は音声タイプライタや自動翻訳機械の実現に必要であるために盛んに研究されてきた。しかし、ラジオで放送された音声テキストの中の単語音声を認識するには、不特定多数の語彙の標準音声パターンを非常に多く記憶して、それらを入力音声パターンと並列照合する必要があるので、登録語彙数だけを考えても経済的な実現が容易ではなかった。更に、音声の登録時と検索時のばらつきの大きさを考えると、従来の音声認識機械の多くは入力音声にかなりの制約を設けていた。それは従来音声認識機械での録音音声パターンと入力音声パターンとの比較照合における柔軟性の不足による。すなわち、入力パターンが分類されるときに、時間的ピッチの揺れはダイナミック・プログラミング(DP)手法によって、ある程度許さ

信号波形の高さを色の濃さで示したものであり、成人男性話者の“島”[s i m a]を発声したときの一例である。垂線204は音素の境界を示している。この音声スペクトログラム203が音声認識の対象となる2次元の音声パターンである。音声認識の基本はこのような音声パターンの照合である。このときに2次元の音声パターンが発声毎に大きく揺らぐ。音声の時間的な揺れが横軸方向のパターンの変動に対応し、音声の高さの揺れが音素単位でのパターンの上下の変動に対応する。強さの揺れは音素毎でのパターンの濃さの変化に対応する。それらを考慮した上で、各入力音声パターンを基準となる複数個の登録音声パターンと一斉に比較することが音声認識技術に求められてきた。

本発明の目的は、各時刻での音声スペクトログラムの上下左右の変動と濃度の変動に伴う音声パターンの認識率の低下を防ぐことにあり、より具体的には、各時刻での音声スペクトログラムの上下左右強弱変動に影響されない音声パターンの認

識技術を提供すると共に、入力音声テキストデータをすべてコード化するのではなく、検索したい文字列の音声スペクトログラムのみを登録することで、その音声文字列を含む音声テキストデータを見つけ出す音声パターンの認識技術を提供することにある。更に、入力音声の時間、高低及び強弱の揺れに左右されない音声照合を可能にするマッチングCCDデバイス技術を提供することにある。

〔課題を解決するための手段〕

本発明の音声テキスト検索装置は、(1) 音声テキスト及び検索用単語音声を周波数別のアナログ信号波形に変換するスペクトル分析手段と、変換された前記アナログ信号波形の各時刻でのアナログ信号振幅を周波数順に光パルス信号に変換する光パルス発生手段と、前記検索用単語音声を重み係数の光マスクパターンに変換して保持すると共に前記光パルス信号を受ける光学マスク手段と、前記光学マスク手段を通過した前記光パルス信号を検出し、前記アナログ信号波形のマッチン

グ処理をし、マッチした単語音声のクラス判別を行うマッチングCCDデバイスとを備えたこと、(2) 前記(1)項に前記マッチングCCDデバイスが前記光学マスク手段を通過した前記光パルス信号を検出し、周波数毎の検出信号を積算して転送するイメージセンサ手段と、前記イメージセンサ手段に結合して各時刻における前記イメージセンサ手段の各段の出力とそれ以前の出力との中から最大値を選び、前記最大値と前段からの転送信号とを比較して最小振幅の方を選択し、次段への転送信号とするミニマクス電荷転送手段と、複数の前記ミニマクス電荷転送手段の出力の中の最大値を検出し、その位置をエンコードする最大検出エンコード手段とを備えたこと、(3) 前記(1)項に前記音声テキストをデジタル信号波形に変換して記憶する音声テキスト記憶手段と、前記検索用単語音声と、前記音声テキスト記憶手段のレコードアドレスと前記マッチングCCDデバイスの出力するクラスコードとでアクセス番地の決まるマッチビット記憶手段とを備えたこ

とを特徴とする。

〔実施例〕

次に図面を参照して本発明を詳細に説明する。

第1図は本発明の音声テキスト検索装置の基本構成を示す説明図である。第1図において、音声テキストや検索用単語音声はまず音声入力手段110であるマイクロホン102によって電気信号に変換され、次にA/D変換器104によってデジタル信号に変換される。そしてスイッチ135を介して音声メモリ120に貯えられる。音声入力手段110は音声信号をデジタル信号に変換するときに音量の正規化やノイズ除去などの平滑化や圧縮などの処理を必要に応じて行う。更に音声メモリ120は再生時のアナログ信号が書き込み時のアナログ音声信号に一致すると共に、周波数スペクトル回路で分析し易くなるように符合化したデジタル信号を格納する。したがって、A/D変換器104は単に各時点でのアナログ値をデジタル値に変換するだけでなく、必要に応じて上述の目的にあった符合化の機能を持っている。音声テキスト

の構成要素となる単語音声を認識するために不可欠なスペクトルアナライザ130は音声入力手段110か又は、音声メモリ120から読み出されるデジタル信号をスイッチ135を介して入力し、高周波にフーリエ変換とかウォルシュ変換等を行う。これにより、音声テキストの各時刻における音声の各周波数成分の振幅がアナログ信号波形として求まる。スイッチ140はスペクトルアナライザ130の出力である周波数別のアナログ信号波形を各時刻で低周波の方から高周波の方へ、あるいは高周波の方から低周波の方へ順に走査し、音声スペクトログラムの各時刻における音声パターンをシリアルに出力し、光パルス発生制御回路150に供給する。またスイッチ140の走査のために周波数走査用カウンタ142と、周波数走査用カウンタ142にクロック信号を供給するクロック発生器144が使われる。光パルス発生制御回路150はレーザダイオードなどの発光ダイオード152が発生する光パルスの強度やパルス幅を制御し、各時刻のアナログ信号波形の振幅を読み取って、その振幅に等

しいパルス幅の光パルス駆動信号を発生する。これにより、周波数別のアナログ信号を、時々刻々とシリアル走査によって光パルスの信号列に置き換える。この光パルスの振幅とパルス幅はCCDのイメージセンサ部182で検出されるに都合のよい大きさに決められる。すなわち、パルス幅を大きくできないときには振幅を大きくし、発光ダイオード152の出力が大きいときにはパルス幅を長くする。ここに、1秒程度の単語音声で60 msec単位でシリアルな光パルス列に変換し、周波数のサンプル点を60とすると、パルス幅は自ずと1 msec以下に制限される。また、発光ダイオード152が10 mW以上であればパルス幅はこの程度で十分である。発光ダイオード152から出力される光パルスは光学レンズ154を通して10～30 mm角程度の光学マスク170に均一に照射され、マッチングCCDデバイス180に突き抜ける。光学マスク170は偏光板と検光板に挟まれたスーパーツイステッド・ネマチック(STN)液晶などの表示マトリックスであって、各メッシュの

光の透過度がそこに印加された電圧によって制御され、設定された透過度がしばらくの間は保持されるものであって、透明な薄膜トランジスタを100万個程度液晶パネル内にマトリクス状に配列し、メッシュの液晶への印加電圧を変えることで、光の透過か遮断かの状態を設定できる。この光学マスク170にはマスクパターンを記憶するマスクメモリ160が接続され、光学マスクにおける重み係数のマスクパターンがマスクメモリ160からの定期的なリフレッシュによって保持される。マスクパターン発生器165は音声メモリ120から読み出される検索用単語音声に対するスペクトルアナライザ130とスイッチ140の処理結果を重み係数列に変換する部分であり、この出力がマスクメモリ160に貯えられる。マスクパターン発生器165は液晶パネルの4×4のメッシュで16値のアナログ量を記憶するようにコード変換を行って重み係数を作る。重み係数列はマスクメモリ160に貯えられた後で、光学マスク170の液晶パネル面上に設定される。1つの単語音声に約1万

6000メッシュを使用するが、薄膜トランジスタは数10単語音声の登録が可能である。マッチングCCDデバイス180はイメージセンサ部182とミニマスク電荷転送部184及び最大値検出型エンコーダ186とを含み、電荷転送制御回路188によって制御される。これは光学マスク170を通過した光パルスを一斉に検出し、どこでマッチングがあったかをシストリックアレイとして働くイメージセンサ部182で調べ、更に音素列のマッチングの有無をミニマスク電荷転送部184で調べてマッチングの結果を最大値検出型エンコーダ186から出力する。イメージセンサ部182はフォトダイオードセルとCCDセルより成るCCDセルアレイであって、光学マスク170を通過した光パルスの透過光量に比例した電荷をフォトダイオードセルに貯え、各フォトダイオードセルに対応したCCDセル上で累積電荷量として転送する。最終段のCCDセルの累積電荷量は各時刻の音素のマッチング結果を示し、次の周波数操作が終るときのみ、ミニマスク電荷転送部184へ転送され

る。これによって、音声の各時刻での周波数スペクトログラムの照合が行われる。ミニマスク電荷転送部184はCCDレジスタとミニマムセクタ及びマキシマムセクタとから成るアレイであって、複数のイメージセンサ部182の各出力段CCDセルの累積電荷量を各段のミニマムセクタで、前段マキシマムセクタから送られてきた列マッチ電荷量と比較し、最小の方を次の段CCDレジスタに送り出す。この処理を各段で並列に行う。マキシマムセクタから送られてきた列マッチ電荷量とは各段から送り出す列マッチ電荷量と前段からの列マッチ電荷量と前前段からの列マッチ電荷量との中の最大値の列マッチ電荷量である。これによってダイナミックプログラミング法による音素列のマッチング処理が行われ、音声の発生速度のばらつきによるミスマッチが救済される。最大値検出型エンコーダ186は多数のミニマスク電荷転送部184の出力の中の最大値を捜し、その位置を単語音声のクラスコードとして出力する部分である。最大値はミニマスク電荷転送部

184 の出力と比較される比較基準の電圧 V をスウィープすることで各時刻に求められる。比較基準の電圧 V は数 10 msec に一度の割りでゆっくりとスウィープされればよい。また比較基準の電圧 V の変化の幅は入力音声の音量に比例するものとするれば、基準値以上の列マッチ電荷量が発生したときにマッチ信号に合わせてクラスコードを出力することにより、単語音声の振幅のゆらぎを意識しないクラス判別を行うことができる。マッチ信号メモリ190 はマッチングCCDデバイス180 からマッチ信号が発生されたときのクラスコードと入力音声テキストを記憶している音声メモリ120 のレコードアドレスとによってマッチ信号の書き込み番地が決まるランダム・アクセス・メモリ(RAM)である。すなわち、アドレス発生器125 からレコードアドレスを音声メモリ120 に与え、レコードアドレスを変えながら音声テキストを読み出して、スペクトルアナライザ130、スイッチ140 及び光パルス発生制御回路150 を通して発光ダイオード152 に送り、そこで光パルス列

に変換する。その光パルス列を光学マスク170 を通してマッチングCCDデバイス180 に入力し、光学マスク170 に登録した検索用単語音声と比較し、マッチング処理によってクラスコード判別を行う。その結果、マッチがあると、そのクラスコードと、そのときの音声メモリのレコードアドレスを用いて、マッチ信号メモリ190 にマッチ信号(例えば、デジタル信号の“1”)を書き込む。マッチ信号メモリ190 の内容は検索用単語音声毎に次のように読み出される。まず、検索用単語音声を音声メモリ120 から再度読み出して、それを光パルス列に変換する。それを同じ単語音声を保持する光学マスク170 に通過させると、マッチングCCDデバイス180 でマッチ信号がすぐに出力される。このマッチ信号に合わせて出力されるクラスコードをマッチ信号メモリ190 に与えると共に、音声メモリ120 と切り離してレコードアドレスのみを高速走査し、マッチ信号メモリ190 の内容を順次に読み出す。そして前述の“1”のマッチ信号が読み出されたときのレコードアドレ

スが求めるマッチレコードアドレスになる。

上述したように、本発明の音声テキスト検索装置は周波数を走査するスイッチ140 の使用と、マッチングCCDデバイス180 のイメージセンサ部182 でのシストリックアレイ法によるベクトル内積演算によって音声の高低のばらつきを除くことができ、マッチングCCDデバイス180 の最大値検出型エンコーダ186 の最大値検出によって音声の強弱のばらつきを除くことができる。また、マッチングCCDデバイス180 の中のミニマスク電荷転送部184 でのダイナミックプログラミング法による音素列のマッチング処理によって発声速度のばらつきを除外できる。

第2図は音声パターンの説明図である。

第3図は音声スペクトログラムの説明図である。第3図は音声スペクトログラムの各時刻での周波数走査の結果であり、第2図の時刻 $t=0.4$ のときの周波数走査による音声スペクトログラムの観測データである。実線は低音の人の、破線は高音の人のデータの一例である。高音の人

の音声スペクトログラムは低音の人の音声スペクトログラムを右へ少しシフトした形になる。また振幅は音声の高さによって異なってくる。第3図により高音や低音を出す人の声道の長さのばらつきや、強さのばらつきが避けられないこと、また第2図から発声速度によって音素領域のばらつきがかなり大きくなることが避けられないことがわかる。音声認識においてはこれらの3種のばらつきに左右されないスペクトログラムの照合が求められる。

第4図は光パルスによるCCDでの電荷発生量とパルス幅との関係を示す説明図である。第4図において、光学マスクの 4×4 メッシュで遮られない ($w=1.0$) とするときの電荷発生量 q は光パルスのパルス幅に比例して増大し、最大値 q_m に達し、そこで飽和する。そのときの光パルスの強さを I_0 とする。パルス幅の変動幅を t_m までとすると、関係特性の実線440 はパルス幅が t_m のときに q が最大値 q_m になるように I を I_0 に設定した場合の一例である。 I を $1.4I_0$

にすると一点鎖線430のように t_m 以下のパルス幅で飽和が起る。また破線450は光学マスクの 4×4 メッシュが重み係数 w に比例して光パルスの通過量が減少する場合を示している。 w が $10/16$ であると、破線450は実線440の場合の $10/16$ の特性となる。パルス幅を最大値 t_m から $x \cdot t_m$ に減少させると、イメージセンサ部182での電荷発生量 q は $w \cdot x \cdot t_m$ に減少する。これがベクトル内積 $\sum w \cdot x$ の積項の演算に相当する。

第5図はイメージセンサ部182での内積演算処理の説明図である。一般にイメージセンサ部182は水平に並ぶフォトダイオードセル510のアレイとCCDセル520のアレイ及びこれに垂直に並ぶCCDレジスタ530とから成る。フォトダイオードセル510が第4図で説明した $w \cdot x \cdot t_m$ の電荷を検出する部分である。通常のイメージセンサではこれが水平の転送ゲートライン515を介してCCDセル520へ移されると、CCDセル520のアレイ上の電荷信号は水平の電荷転送クロック線

525からのシフトクロックパルスを受けて一斉に右へシフトされる。CCDレジスタ530に到達した電荷は垂直の転送ゲートライン516を介して垂直の電荷転送クロック線535のシフトクロックパルスを受けて一斉に下へシフトされる。本発明においては、ここで信号波形のマッチングを行う。そのためには、電荷転送クロック線525からシフトクロックパルスを1回印加する都度、転送ゲートライン515を1回駆動し、フォトダイオードセル510の電荷をCCDセル520で転送される電荷に次々と加算していく。

数学的に言えば、 j 行 i 番目のフォトダイオードセル510の電荷量 $q_{ji} = [w_{ji} x(t)]$ を j 行の i 番目のCCDセル520に移した後、電荷転送クロック線に1クロック加え、それは j 行の $(i+1)$ 番目のCCDセル520に残る。1セル分の電荷をシフトした後で、 $x(t+1)$ の光パルスに対する入射光量が $(i+1)$ 番目のフォトダイオードセル510に貯えられる。そのときの $w_{ji+1} x(t+1)$ の電荷を j 行 i 番目のCCDセル520に移すと、

そこで、 $q_{ji} = w_{ji} x(t)$ との加算が起る。すなわち、次式の電荷の加算結果、

$$q_{ji+1}(t+1) = q_{ji}(t) + w_{ji+1} x(t+1) \quad (1)$$

が $(i+1)$ 番目のCCDセル520に残る。また、 j 行 i 番目のCCDセル520では同じ時刻に次式の電荷の加算結果は、

$$q_{ji}(t+1) = q_{ji-1}(t) + w_{ji} x(t+1) \quad (2)$$

が貯えられている。 i が1から n まで変わるとすると、 n 番目のCCDセル520では、次の漸化式で表わされる電荷は、

$$\begin{aligned} q_{jn}(t+1) &= q_{j1-1}(t) + w_{jn} x(t+1) \\ &= q_{j0} + \sum_{i=1}^n w_{ji} x(t+1) \end{aligned} \quad (3)$$

が求まる。ここに $t+1$ は t に置き換えてもよく、光パルスの印加回数である。 n 番目のCCDセル520に到達した累積電荷は転送ゲートライン516を介してCCDレジスタ530へ移される。このとき、上のCCDレジスタ530から転送された電荷とCCDレジスタ530上で加算される。また、水平方向からの n クロック転送毎に1回の垂

直クロックを電荷転送クロック線535から受けて、加算結果の累積電荷を下のCCDレジスタ530へ送ることが可能である。したがって、最下段のCCDレジスタ530には常に段数に等しい期間前から各時刻までの音声の照合結果を表わす累積電荷が到達する。これをそのまま最大値検出型エンコーダ186に送ることで音声単語の検出が可能である。しかし、これでは音声の時間的な揺れに弱くダイナミックプログラミング法によるマッチング処理を必要とする。そのため、第5図の垂直のCCDレジスタアレイはこのままでは音声の照合には使用できないので、代りに第6図のミニマスク電荷転送部184を使用する。この場合、 t は n 個の周波数を下から上に順に走査したときのサンプル番号に対応する。なお、各行のCCDセルアレイでは、シストリックアレイ法に基づく漸化式の演算によって重み係数列と光パルスの列とのベクトル内積が計算され、その結果が最終段のCCDセル520から求められ、それが周波数走査の都度出力される。また、どの周波数でマッ

ングのピークがあっても、それが最終段のCCDセル520まで転送されるので、ピーク周波数の揺れに強いマッチング処理が達成される。

第6図はミニマスク電荷転送部184と最大値検出型エンコーダ186でのダイナミック・プログラミング・マッチング動作を説明する図である。第6図において、CCDセルアレイ500はイメージセンサ部182の各行を代表して示している。また、各行のCCDセルアレイ500は入力音声スペクトログラムと登録音声スペクトログラムの各時刻における音声スペクトログラムの照合結果(類似度)を電荷量の周波数走査の終了の都度出力する。ミニマスク電荷転送部184はj列目のCCDセルアレイ500で求まる電荷 q_{mj} を隣のCCDレジスタ610に転送するとき、隣のCCDレジスタ610から転送されてくる信号 $z_{j-1}(t-1)$ と加算するのではなく、小さい方を選んで送る。それを列マッチ信号とすると、それは、

$$z_j(t) = \min[z_{j-1}(t-1), q_{mj}(t+1)], \\ j=1, 2, \dots, m \quad (4)$$

ると、そのときには $z_j(t-1)$ がピークを迎えるので、それを $z_j(t-1)$ の代りに使えばよい。したがって、ミニマスク電荷転送部184はCCDレジスタ610とミニマムセクタ620とマキシマムセクタ630を繰り返し配列したものとなり、その処理機能は、

$$z(t) = \min[q(t), \max[z_{j-2}(t-1), \\ z_{j-1}(t-1), z_j(t-1)]] \quad (6)$$

で表わされる。処理結果は常にCCDレジスタ610に保持される。これによって、入力音素列が登録音素列よりも早くなっても遅くなっても、音素片列の類似度を示す列マッチ信号 $z_j(t)$ が急速に減少しないようになる。最大値検出型エンコーダ186は最大値検出回路640によって列マッチ信号と比較される基準の電圧Vを高い値から低い値に減少させたときに、最初に基準電圧を越える列マッチ信号を発生したミニマスク電荷転送部184の位置を検出し、エンコーダ650でマッチ信号に合わせてクラスコードを出力する。なお、基準電圧Vは入力音声の音量に比例した電圧値迄しか下

であって、ミニマムセクタ620から出力される。この出力を次の段に送って同様にこれを繰り返すだけでは、

$$z_m(t) = \min[q_m(t), q_{m-1}(t-1), \dots, \\ q_0(t-m)] \quad (5)$$

となって、各時刻の光パルス列と重み係数列とがタイミングよく次々と一致すれば大きな類似度を示すが、タイミングに揺れがあると小さくなってしまふ。この部分にダイナミック・プログラミングによるマッチング手法を導入すると、発音速度の揺らぎによるミスマッチを防ぐことができる。そのためには、(4)式の右括弧の中の $z_{j-1}(t-1)$ の代りに、

$$\max[z_{j-2}(t-1), z_{j-1}(t-1), z_j(t-1)]$$

を用いればよい。すなわち、タイミングが合っていれば、 $q_j(t)$ がピークになる時刻に $z_{j-2}(t-1)$ がピークになる筈であるが、 $q_j(t)$ が早目にピークになると、 $z_{j-2}(t-1)$ がピークを迎えているので、それを $z_{j-1}(t-1)$ の代りに使うと、時間補正がうまくいく。また、 $q_j(t)$ が遅目にピークにな

らないため、登録音声にマッチした音声が入力されない場合にはマッチ信号やクラスコードは出力されない。

第7図は重み係数を保持する光学マスクの実施例の説明図である。光学マスクは光パルス列の透過を制御して、検索用音声とテキストの中の音声との各時刻の各周波数でのスペクトログラムの比較を行う部分であって、本発明の中で重要な役目をする。第7図の例では、光学マスク170は電極板720を付着した偏光板740とアモルファス薄膜トランジスタ(TFT)のアクティブマトリクス730を付着した検光板750との間にツイステッド・ネマチック(TN)とかスーパー・ツイステッド・ネマチック(STN)とかの液晶710を挿入したものである。アクティブマトリクス730は600×600程度の液晶セルに電圧を選択的に印加する。液晶セルの光パルスの透過度は印加電圧を変えることで変化し、概して高電圧のときに透過度が下り、低電圧のときに上がる。この電圧はマスクメモリ160から与えられる。

第8図は光学マスクのアクティブマトリクス駆動回路の説明図である。重み係数のデータによって電圧が変わるソース線810と液晶セル840の位置を電圧によって順次選択するゲート線820との交点部にアモルファス薄膜トランジスタ830がある。ソース線810とゲート線820の電圧が共に高いときにアモルファス薄膜トランジスタ830はオンになり、ドレイン電圧が高くなって、それと電極850との間に接続される液晶セル840が透明から不透明に切り替わる。ドレイン電圧が決まった後では、ゲート線820の電圧が下ってもドレイン電圧がしばらくそのままに保たれる。したがって、個々の液晶セル840に個別に駆動電圧を印加しなくても光学マスクに登録した重み係数のマスクパターンは消滅しない。ただし、アモルファス薄膜トランジスタ830でのオフ電流が少しずつドレイン電圧を低下させるので、数10ヘルツ程度でのリフレッシュがマスクパターンの保持には必要であり、このためにマスクメモリ160が使用される。

らでも音声の照合を行えること、(b) 音声の高さ、及び低さに影響されず、各時刻での周波数スペクトログラムのマッチング結果を出力できること、(c) 音声の発生速度の揺れに左右されずに、音素列のマッチング処理を行えること、(d) 音声の強さに左右されずに音素列のマッチング結果を比較できること等が可能になる。このような音声の3つの揺れに対して柔軟性を持たせたことの効果は検索漏れを少なくできるところと、それを無くするために必要な検索用単語音声の登録数を減らせることにある。例えば、音声テキストでの単語音声の高さが5種類に及び、強さも5種類に及ぶと、登録する単語音声数が25になるが、それを1つの検索用単語音声でカバーできるとすれば、重み係数を保持する光学マスクが1/25で済むことになる。また、音声テキストの検索を音声テキストすべてを認識してから行うのではなく、直接、音声テキストから検索用単語音声を見つけ出すように構成を工夫したので、これによって、検索用単語音声数が数10個になっても、

〔発明の効果〕

以上詳細に説明したように、本発明の音声テキスト検索装置によれば、従来の音声認識機械での入力音声の時間、高低及び強弱の揺れに強くない音声認識方式の問題点を解決するために、(A) スペクトルアナライザで得た音声の周波数別のアナログ信号波形を各時刻で周波数順に走査して光パルス列に変換し、すべての登録音声のすべての周波数のアナログ信号波形とすべての時刻の成分とを比較できるように、光パルスを光学マスク全面に照射するようにし、(B) 光学マスクを通過した光パルスを各時刻でCCDデバイス上で集計(シストリックアレイによるベクトルの内積演算を実行)した。(C) その結果をマッチングCCDデバイス上でダイナミックプログラミング法で処理することにより入力音声と各登録音声との類似度を求め、最大値検出型エンコーダでクラス判別をするようにした。

上述のように音声認識を3段階に分けたことで、(a) 音声の始まりを認識しないで、いつか

一度の音声テキストのサーチで検索結果が求まる。これは検索時間が1/10に減少することに相当する。また、全部の音声テキストを音声認識でコード化する必要が無く、直接に単語音節で検索できるので、ラジオやテレビから得た音声テキストのデータベース化の工数が大幅に低減されるという効果がある。更に、本発明の新しい音声認識方式を液晶光学マスクとCCDデバイスで実現する方法を提示した。これらは容易に一体化されるので十分小型化の実現も可能であって携帯に適した大きさになり、例えば国際会議の会場に携帯し会議場で口頭で発表される論文をマイクロホンでモニターし、その場で口頭の音声テキストの内容検索を行うことが可能になる等の効果がある。

図面の簡単な説明

第1図は本発明の一実施例の基本構成を示す説明図、第2図は音声パターンの説明図、第3図は音声スペクトログラムの説明図、第4図は光パルスによるCCDでの電荷発生量とパルス幅との関

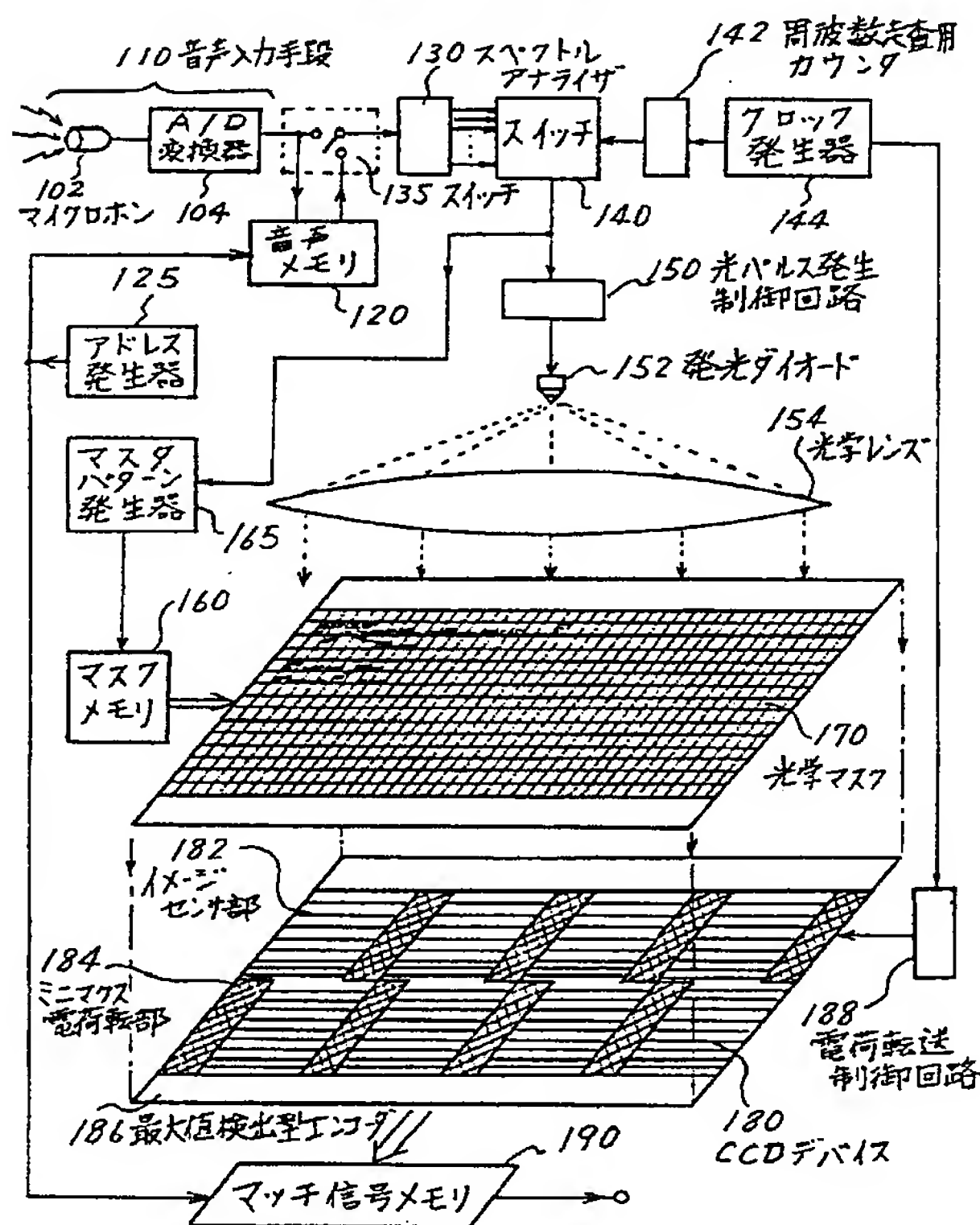
係の説明図、第5図はイメージセンサ部での内積演算処理の説明図、第6図はミニマスク電荷転送部でのダイナミックプログラミング・マッチング動作の説明図、第7図は光学マスクの一実施例の説明図、及び第8図は光学マスクのアクティブマトリクス駆動回路の説明図である。

102 … マイクロホン、104 … A/D変換器、110 … 音声入力手段、120 … 音声メモリ、125 … アドレス発生器、130 … スペクトルアナライザ、135、140 … スイッチ、142 … 周波数走査用カウンタ、144 … クロック発生器、150 … 光パルス発生制御回路、152 … 発光ダイオード、154 … 光学レンズ、160 … マスクメモリ、165 … マスクパターン発生器、170 … 光学マスク、180 … マッチングCCDデバイス、182 … イメージセンサ部、184 … ミニマスク電荷転送部、186 … 最大値検出型エンコーダ、188 … 電荷転送制御回路、190 … マッチ信号メモリ、201 … 時間軸、202 … 周波数軸、203 … 音声スペクトログラム、204 … 垂線、430 … 一点鎖線、440 … 実線、450 … 破線、500 …

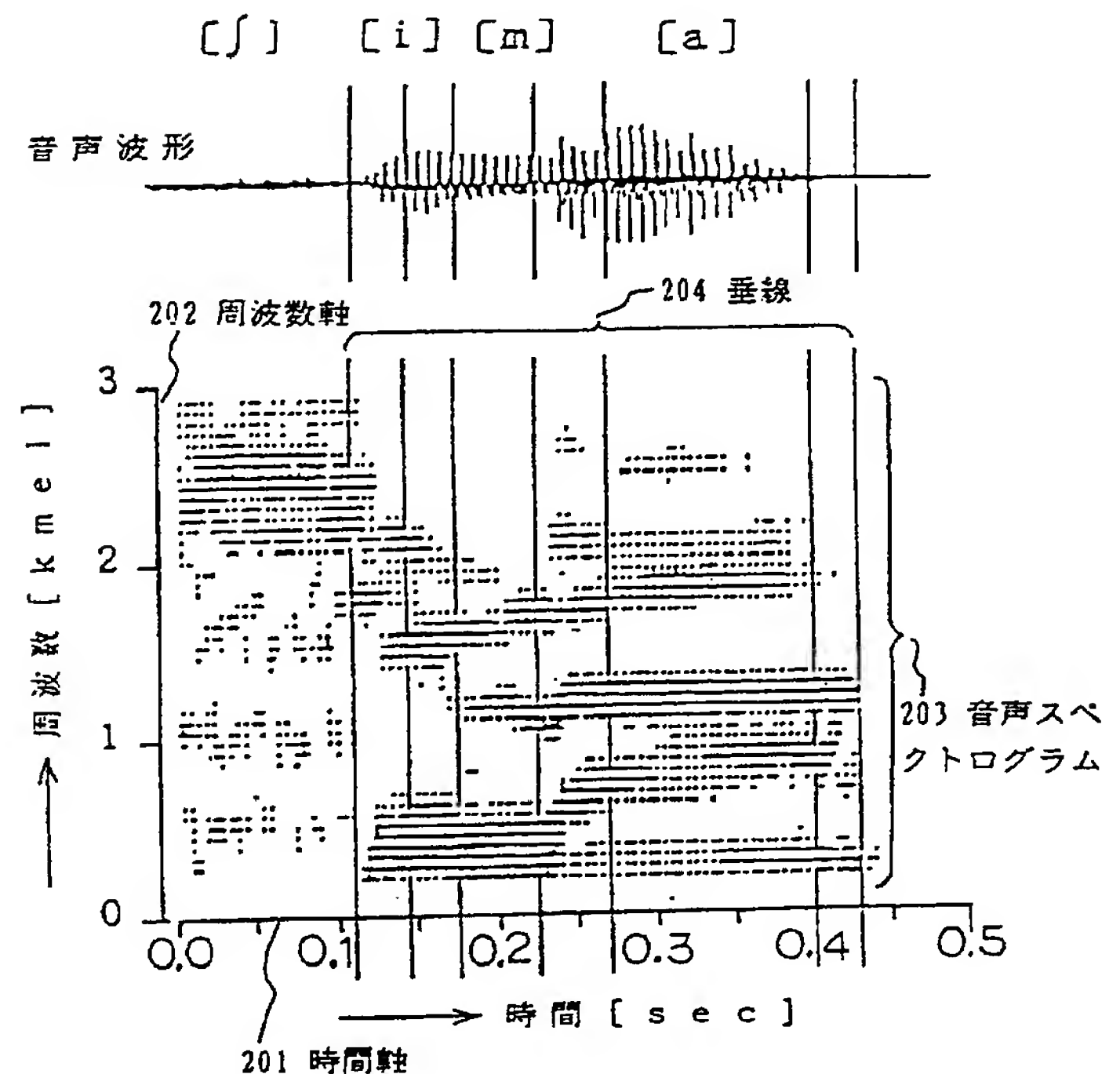
CCDセルアレイ、510 … フォトダイオードセル、515、516 … 転送ゲートライン、520 … CCDセル、525、535 … 電荷転送クロック線、530、610 … CCDレジスタ、620 … ミニマムセクタ、630 … マキシマムセクタ、640 … 最大値検出回路、650 … エンコーダ、710 … 液晶、720 … 電極板、730 … アクティブマトリクス、740 … 偏光板、750 … 検光板、810 … ソース線、820 … ゲート線、830 … アモルファス薄膜トランジスタ、840 … 液晶セル、850 … 電極。

代理人 弁理士 内 原 晋

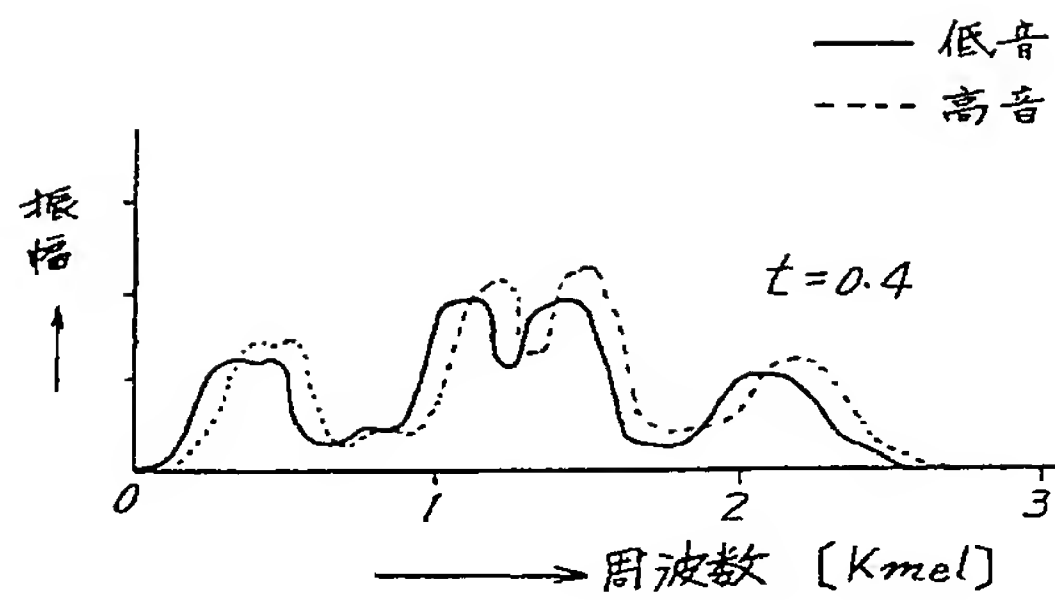
第 1 図



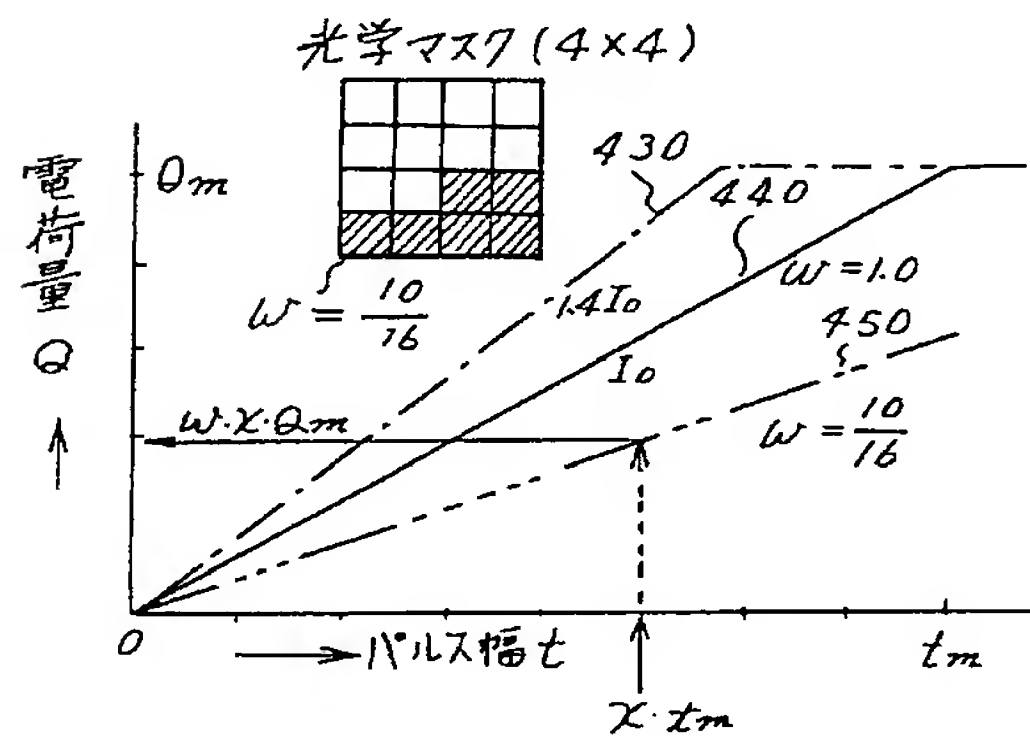
第 2 図



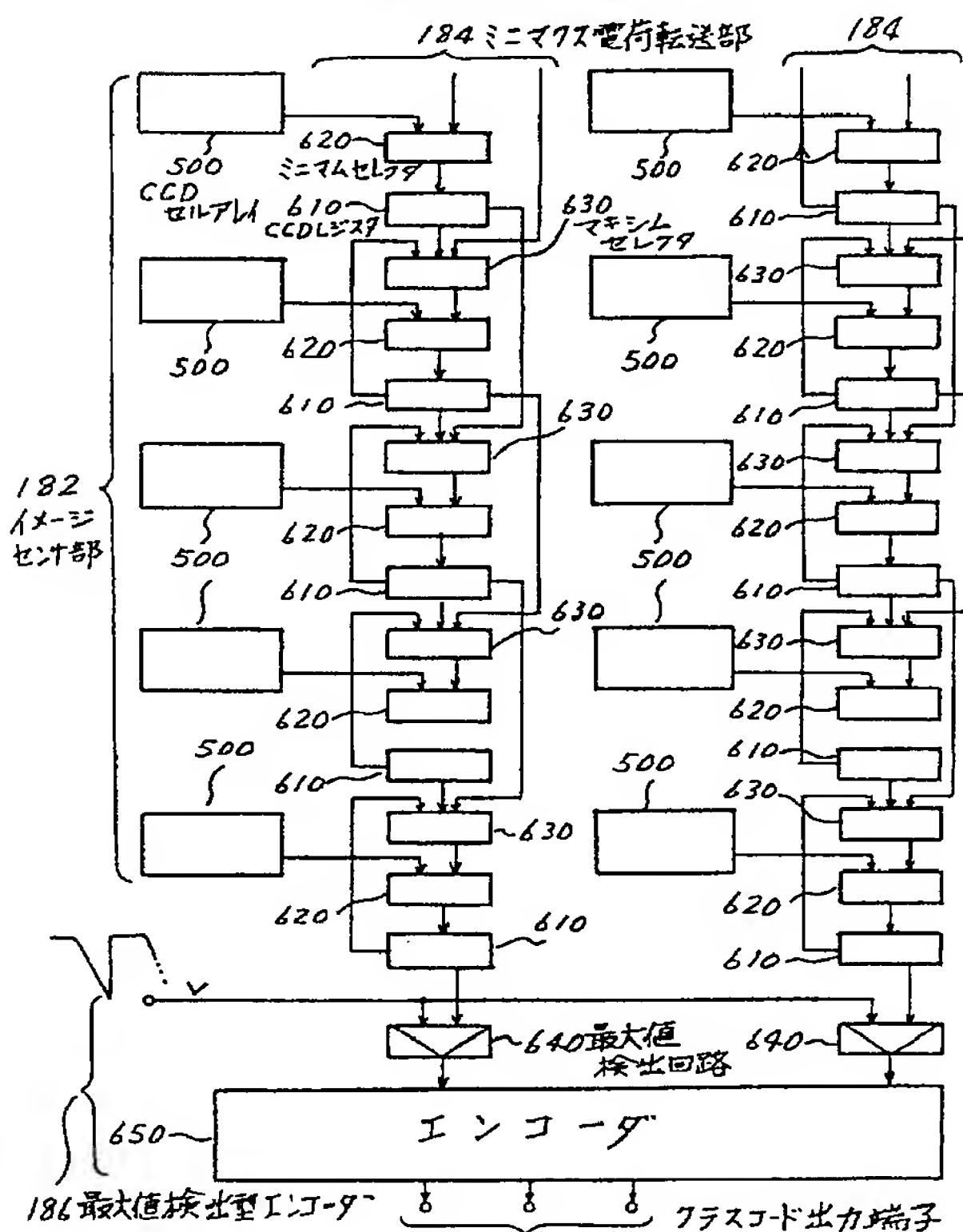
第 3 図



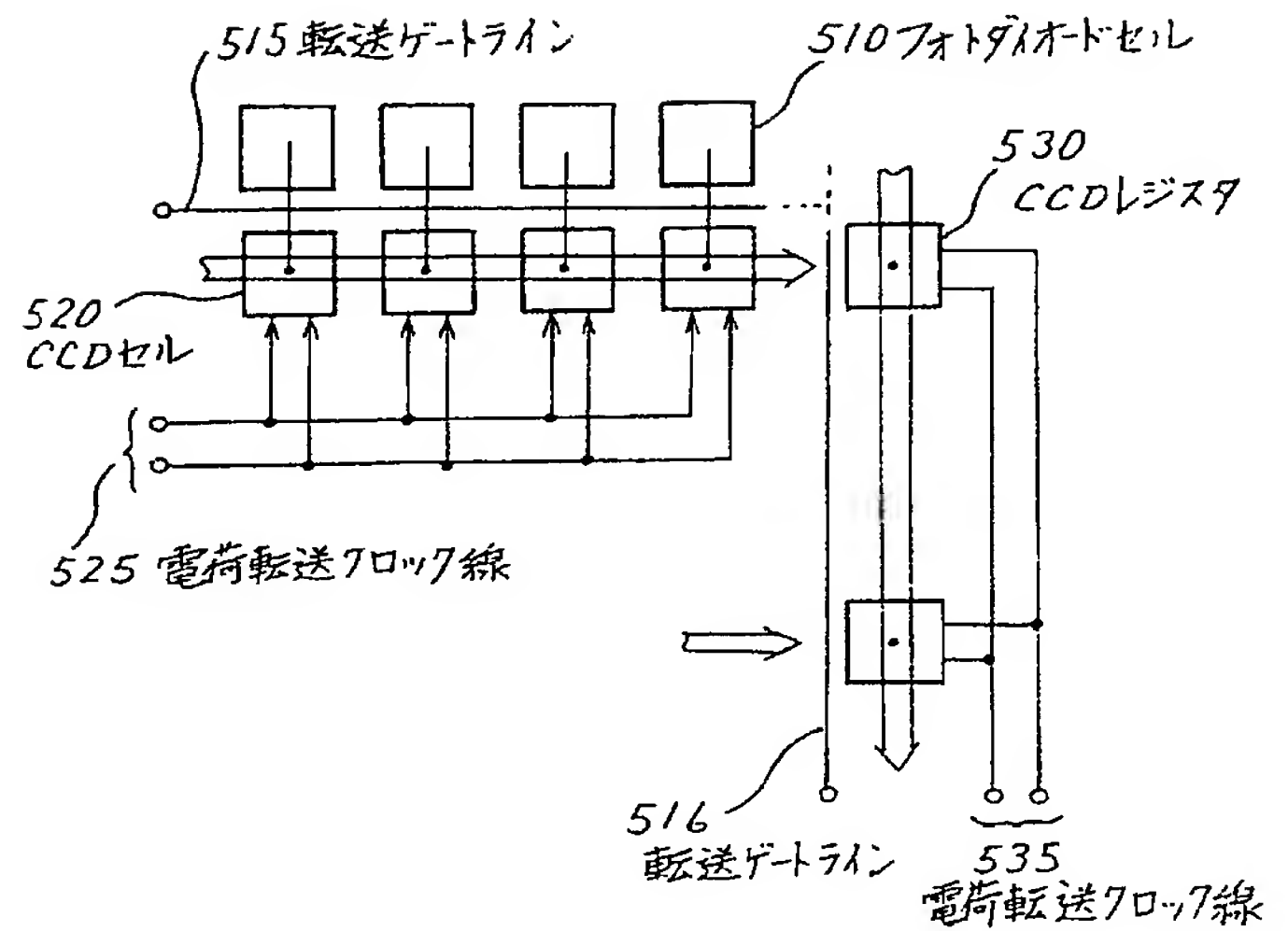
第 4 図



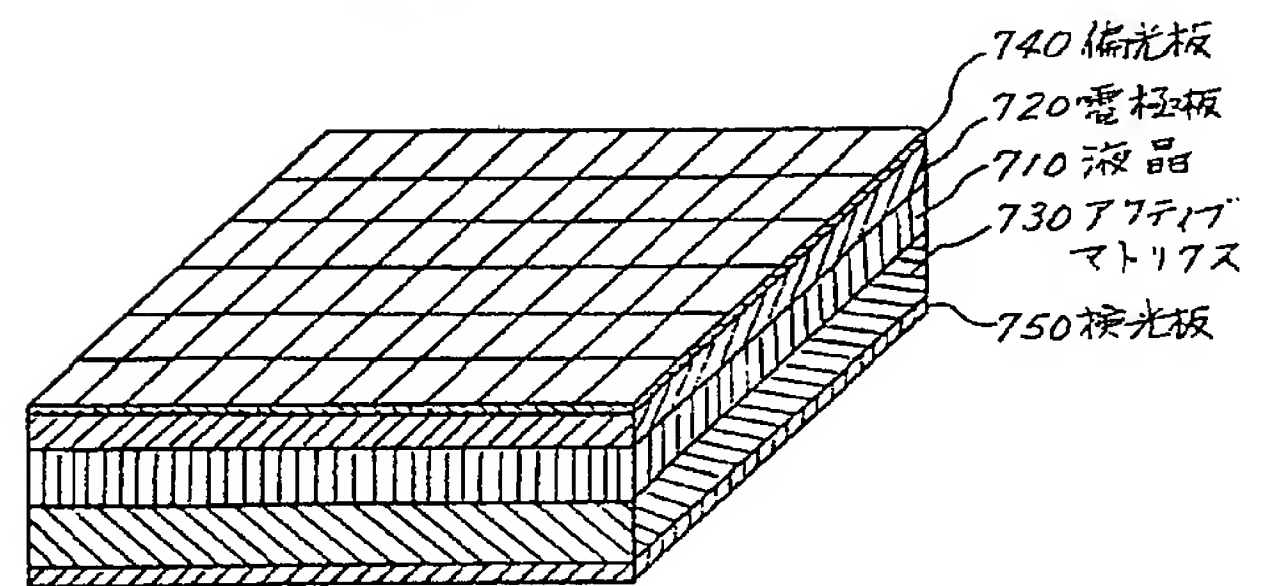
第 6 図



第 5 図



第 7 図



第 8 図

